

December 23, 2024

E-Filed

The Honorable Thomas S. Hixson
United States District Court for the Northern District of California
San Francisco Courthouse, Courtroom E – 15th Floor
450 Golden Gate Avenue
San Francisco, CA 94102

Re: *Kadrey et al. v. Meta Platforms, Inc.*; Case No. 3:23-cv-03417-VC-TSH

Dear Judge Hixson:

The parties jointly submit this supplemental letter brief concerning issues related to Plaintiffs' RFP No. 118 addressed during the December 19, 2024 discovery hearing before Your Honor.

I. PLAINTIFFS' STATEMENT

At the December 19 hearing on, *inter alia*, Dkt. No. 308, the Court ordered Plaintiffs identify in a supplemental letter brief the specific and narrowed post-training data they're seeking via RFP 118. *See* Dkt. No. 308 at 1 (asking Court to order production of data referenced in Meta's Llama 2 and 3 papers).

A. Relevant Background

Meta produced hard drives containing a single copy of books datasets it copied from Books3 and Library Genesis and used for “pre-training” Llamas 1-3.¹ Both of these pirated books datasets contain Plaintiffs' Asserted Works. Plaintiffs currently seek the *post*-training datasets Meta used to study Llama's tendency to memorize copyrighted data on which it has trained, and to prevent its models from regurgitating this copyrighted data when prompted by end users. These post-training datasets were also copied from Shadow Libraries containing Plaintiffs' Asserted Works and thus are directly relevant to Plaintiffs' copyright infringement claim because they reflect Meta's copying and varied uses of Plaintiffs' Asserted Works to train its Llama models.

The parties dispute the conceptual relevance of this data, which has been fully briefed at Dkt. No. 308. To briefly reiterate, this post-training data is relevant for at least the following reasons:

- Plaintiffs' copyright infringement claim is predicated on Meta's copying and uses of Plaintiffs' Asserted Works for “training” its Llama models. Dkt. No. 133 at ¶ 79. The “full training process”—as Meta's own corporate deponent testified—comprises both “pre-training” and “post-training” phases. Ex. A at 54. Meta deponents also testified that copyrighted books sourced from Shadow Libraries containing Plaintiffs' Asserted Works were used in the post-training phase. *See, e.g.*, Ex. B. Thus, Meta's post-training datasets comprising copyrighted books are evidence of Meta's copying and myriad uses of Plaintiffs' Asserted Works in training its Llama models.
- Meta relies on its post-training mitigations as support for its fair use defense. *See, e.g.*, Dkt. No. 336-1 at 7 (ROG response describing Meta's “efforts to minimize [its] models' ability to memorize and/or output training data verbatim” as central to claim that use of Plaintiffs' Asserted Works is “transformative” and therefore “fair use”). These mitigation efforts occur during both the pre-training and post-training phases of model development and necessarily rely on copyrighted books datasets. *See, e.g.*, Ex. B.
- Meta uses copyrighted books datasets to teach Llama to identify language that is copyrighted and to properly respond to requests for outputs of infringing material. These datasets are, therefore, relevant to the question whether Meta's Llama models retain a copy or “memory” of the copyrighted training data on which they train. If they do, then Meta's intended reliance on the “intermediate copying” cases in its fair use defense is unavailing.

¹ Despite the Court's orders that Llamas 4 and 5 are relevant, Meta has not identified the books datasets (including copies) that include Plaintiffs' works it downloaded from *other* shadow databases (e.g., Z-Lib) for training Llamas 4 and/or 5. Meta already knows the specified storage locations where copies exist, *see* Ex. C at 9, yet Meta failed to supplement its response to ROG 1, which explicitly asks for this and related information. Plaintiffs seek this information in letter briefing on Meta's ROG responses filed December 20.

B. What Plaintiffs Seek

At the time Plaintiffs filed the letter brief at Dkt. No. 308, they only had a partial understanding of what Meta possesses in the way of post-training data. Thus, Plaintiffs referred conceptually to sections of the Llama 2 and 3 papers. Those papers functioned as a guidepost for the pre- and post-training data that Meta possess. After Dkt. No. 308 was filed, Plaintiffs deposed Nikolay Bashlykov, a data engineer who testified regarding the actual location, contents, and format of Meta’s post-training data. Plaintiffs also received an additional 3.8TB of source code from Meta, which includes source code regarding memorization mitigations. While Plaintiffs’ experts are still sifting through that code, it has provided further insight into Meta’s specific post-training data repositories, and Plaintiffs’ experts have made clear that their ability to actually understand this relevant source code requires the post-training data sets that Meta uses for testing.

Armed with this new information, Plaintiffs ask the Court to order Meta to produce the following post-training data, which falls under the umbrella term “Supervised Fine-Tuning”:

1. Datasets Identified By Mr. Bashlykov

Plaintiffs filed a letter brief on December 20 that, in part, addressed a set of “supervised fine-tuning data” that Mr. Bashlykov testified is located on a specific hard drive cluster referred to as EAG-WSF. Dkt. No. 356. That brief explains its relevance. This data is also relevant to this supplemental brief. Plaintiffs believe the supervised fine-tuning data identified by Mr. Bashlykov corresponds, at least in part, to § 3.1 of the Llama 2 paper and §§ 4.2.2 and 5.4.7 of the Llama 3 paper.

2. Datasets Identified By Plaintiffs’ Experts Via New Source Code

In reviewing Meta’s recent source code production, Plaintiffs’ experts identified the following post-training datasets as highly relevant discovery. The datasets are referenced within the source code itself but have *not* been produced by Meta. These are not new requests; they are more targeted “asks” under RFP 118 that narrow the broader data from the Llama papers sought in Dkt. 308.

- The post-training datasets used to train and fine-tune the Llama models specifically in reference to the “Intellectual Property” safety category, which is commonly referred to in the source code as “CRS 17 - Intellectual Property.” This type of training is variously referred to as “fine-tuning,” “safety fine-tuning,” “mitigation training,” or other similar terms. Plaintiffs believe this data likely corresponds to § 5.4.7 of Meta’s Llama 3 Paper.
- The post-training datasets comprising books sourced from the at-issue shadow datasets that are used for other safety and CRS categories. This data corresponds, at least in part, to § 4.2 of the Llama 2 paper and §§ 5.4.3 and 5.4.7 of the Llama 3 paper. Plaintiffs are *not* requesting all safety-related datasets; *only* those containing *books* from Shadow Libraries.
- Any additional post-training datasets sourced from shadow datasets and used by Meta to fine-tune its Llama models to minimize their ability to memorize or output training data verbatim. This corresponds to § 3.1 of the Llama 2 paper and §§ 4.2.2 and 5.4.7 of the Llama 3 paper.

Plaintiffs seek both the raw/original data from which these post-training datasets were created, as well as the data as specifically formed or constituted for use in the aforementioned post-training of the Llama models.

II. META'S STATEMENT

The Court should deny Plaintiffs' motion for multiple reasons: Plaintiffs have disregarded the Court's order (Dkt. 351) by failing to narrow the relief requested in Dkt. 308; the requested materials do not fall under RFP 118; and Plaintiffs' requested relief is unwarranted, including because it calls for vague information that is of minimal relevance and would disproportionately burden Meta—particularly given the belated nature of Plaintiffs' request.

A. Plaintiffs' Revisionist Reframing of Dkt. 308 Should be Disregarded.

Plaintiffs' brief is not, in fact, a supplemental brief regarding Dkt. 308 but an entirely new, time-barred request for broader relief: amorphous "Supervised Fine-Tuning" ("SFT") data, which Plaintiffs loosely characterize as post-training datasets, as well as "the raw/original data from which these post-training datasets were created." This approach disregards the Court's specific directive following the December 19, 2024 hearing: to identify the "particular data referred to" in certain sections of the Llama 2 and 3 papers given that at the hearing "Plaintiffs made clear they want certain particular data referred to in those sections, not all of it." Dkt. 351 at 1. Notably, Plaintiffs' new request simply posits, without explanation, that the data they now seek "corresponds" or "likely corresponds" to the Llama papers at issue in Dkt. 308. But the cited Llama paper sections contradict Plaintiffs' characterizations of them—the descriptions of SFT data in the papers have nothing to do with books.²

To justify their deviation from the Court's directive, Plaintiffs allege that after they filed Dkt. 308 on December 4, they received allegedly "new information"—namely, the deposition of Mr. Bashlykov and an additional production of source code. They also seek to minimize Dkt. 308's reliance on the Llama papers—papers that the Court took the time to parse, but that Plaintiffs were unprepared to discuss at the hearing—claiming, falsely, that Dkt. 308 merely "referred conceptually to sections of the Llama 2 and 3 papers," which were merely "a guidepost" for their requests. This argument is belied by Dkt. 308 and the Llama papers, and should be rejected.

Mr. Bashlykov was deposed on December 5 and 6—approximately two weeks *before* the December 19 hearing and the December 20 deadline for filing a discovery motion. If his testimony were relevant to the present motion, Plaintiffs could have and should have told the Court sooner, including at the December 19 hearing. Instead, as Plaintiffs' "supplemental brief" acknowledges, one of their multiple December 20 motions to compel separately seeks the information. *See* Dkt. 356 at 1–2 (seeking "Supervised Fine-Tuning Data"). As set forth in Meta's opposition to Dkt. 356, that request should be denied, including because RFP 118 "is directed to documents describing or showing 'efforts, attempts, and measures,'" regarding mitigation and "does not ask for training data of any kind." Dkt. 356 at 3–4. In all events, Mr. Bashlykov's testimony does not refer to a subset of the data described in the Llama papers and thus has no bearing on this brief.

The source code production referred to by Plaintiffs occurred on **December 2**—before they filed Dkt. 308, not after. *See* Dkt. 308 at 3 (Plaintiffs, acknowledging production *prior to* Dkt.

² *See* Llama 2 § 3.1 (*see* Table 5, an example of SFT data), § 4.2 (distinguishing between different types of safety finetuning, only one of which is SFT and describing use of "adversarial prompts and safe demonstrations," **not** books); Llama 3 § 4.2.2 (describing data "largely comprised of... prompts from our human annotation collection with rejection-sampled responses... [and] Synthetic data targeting specific capabilities"), § 5.4.3 (describing use of data from Llama 2 SFT and a like dataset of prompts and demonstrations with "borderline" examples), § 5.4.7 (referring to similar "prompt and response" examples).

308’s filing). And, again, that production has no bearing on this brief, as it does not involve narrowing the universe of data referred to in the Llama papers to a more discrete set of allegedly relevant information—*i.e.*, what the Court ordered Plaintiffs to brief. Instead, Plaintiffs use the source code produced *prior to their filing of Dkt. 308* as a claimed basis for broadening their ask to include “the raw/original data from which these post-training datasets were created.”

Because Plaintiffs now seek broader and different relief from that sought in Dkt. 308, rather than narrowing the datasets cited in the Llama papers as ordered, their request should be denied.

B. RFP 118 Does Not Cover the “Supervised Fine-Tuning Data Now Requested.”

Critical to a resolution of Dkt. 308 is the actual text of RFP 118, which does not seek *any* training data, including the “Supervised Fine-Tuning Data” that Plaintiffs now request. To the contrary, RFP 118 seeks: “All Documents and Communications, including source code, relating to any efforts, attempts, or measures implemented by Meta to prevent Llama Models from emitting or outputting copyrighted material.” Ex. A to Dkt. 308. As this RFP calls for, Meta has produced documents showing “efforts, attempts, or measures” to prevent data memorization.

Notably, training data was the subject of *other, earlier* document requests by Plaintiffs—namely, RFPs 1–3, served nearly a year ago on **December 27, 2023**, which sought “The Training Data” for Llama 1–3. *See* Ex. D at 10. Those year-old requests defined “Training Data” to include the very information Plaintiffs now seek: “textual or other material used, or considered for use, as input to a machine-learning model... *during any phase* of its development, including pretraining, training, validation, testing, *fine-tuning*, and alignment.” *Id.* at 4 (emphasis added). Plaintiffs did file a motion to compel on RFPs 1-3 – but did not request post-training data – which the Court denied on November 25. *See* Dkt. 288. Any further motion regarding these requests is untimely. *See* Dkt. 253 (confirming October 23 deadline for motions regarding “existing written discovery”).

This Court has made clear that the parties are not at liberty to rewrite their RFPs in the course of briefing motions to compel, and that relief will only be ordered if an RFP actually seeks what a party moves for. *See, e.g.*, Dkt. 351 (denying request for “things not requested by this RFP”). The Court should follow that approach here. RFP 118 does not seek “Supervised Fine-Tuning” data and thus it does not and cannot support Plaintiffs’ requested relief.

C. Plaintiffs’ Requested Relief Is Unwarranted.

Notwithstanding, Plaintiffs’ request for all “Supervised Fine-Tuning” data is unwarranted. The vague information sought is of marginal relevance at most and would impose a hugely disproportionate burden on Meta to locate, collect, and produce. Plaintiffs have all the information they need for their case without wading into this burdensome and tangential territory so late.

As an initial matter, Plaintiffs’ current requested relief, in addition to being different from and broader than what they sought in Dkt. 308, remains highly vague. It thus would require significant effort by Meta to even determine what it is Plaintiffs seek, let alone actually collect, review and produce the information. For example, two of the three categories described by Plaintiffs (*supra* at 2) involve the term “shadow datasets,” and as this Court has observed, that term is in the eye of the beholder. *See* Dkt. 315 at 7–8. Plaintiffs also fail to explain what they mean by “datasets that are used for other safety and CRS categories,” let alone how those have any relevance to the issues presented by their sole claim in this lawsuit.

Plaintiffs also claim that their supposed relevance arguments were “fully briefed at Dkt. No. 308,” but then provide new relevance arguments that they have not made previously. These

belated arguments lack merit, exceed Dkt. 308, and do not justify the burden of the requested relief.

First, Plaintiffs claim that Supervised Fine-Tuning data is relevant as “evidence of Meta’s copying and myriad uses of Plaintiffs’ Asserted Works in training its Llama models.” But Plaintiffs already have information about whether there was or was not copying of their works—namely, the datasets Meta used to pretrain its Llama models. Just as Plaintiffs do not need every copy of their works that Meta might have to establish their claims, *see* Dkt. 351 at 2, they don’t need Meta’s “Supervised Fine-Tuning” data to determine whether copying has or has not occurred. This is especially so given that the data Plaintiffs are seeking, if it exists, would be a subset of the pre-training data Meta already produced. Further, Plaintiffs do not contend, let alone present evidence, that any of the Asserted Works (as opposed to other unidentified copyrighted works in the so-called “Shadow Libraries”) were actually used in fine-tuning any of the Llama models.

Second, Plaintiffs claim that Meta’s fair use defense refers to Meta’s efforts to mitigate against verbatim text output, and that Supervised Fine-Tuning data is relevant to this process. As Meta has previously explained, *see* Dkt. 308, Plaintiffs assert no copyright claim in the case based on alleged verbatim text output. Nevertheless, Meta already has provided extensive discovery and testimony about its mitigations generally, and, again, if Plaintiffs seek to show copying through this information (*see supra*, arguing that post-training “necessarily rel[ies] on copyrighted books datasets”), they already have information about copying from the pretraining phase.

Third, Plaintiffs claim that Supervised Fine-Tuning Data is relevant to “whether Meta’s Llama models retain a copy or ‘memory’ of the copyrighted training data.” This speculative argument is baseless. Meta’s use of fine-tuning datasets does not bear on whether the model “retains” a “copy” of Plaintiffs’ asserted works, and, in any case, Plaintiffs already otherwise have Meta’s source code and taken discovery on that topic. To be sure, the Court dismissed any claim in the case that Meta’s models are a “copy” of Plaintiffs’ works. Dkt. 56.

III. PLAINTIFFS’ REPLY

Relevance. Section 5.4.7 of the Llama 3 Paper expressly states that Meta implements “Intellectual Property” safety measures. *See also* Llama 3 Paper at 51 (chart showing reduction of “Intellectual Property Violation Rate” by nearly 100%). For its part, RFP 118 covers “efforts, attempts, or measures implemented by Meta to prevent Llama Models from emitting or outputting copyrighted material.” That RFP 118 doesn’t expressly mention “supervised fine tuning” is a red herring. This data is squarely responsive because Meta’s copyright mitigation efforts are premised on both source code and associated data. For instance, the memorization source code repository Meta produced earlier this month references some of these SFT datasets. Ultimately, Meta’s SFT process uses copyrighted works (a separate act of infringement) to train Llama on what **not** to emit. Meta cites this as central to fair use. Dkt. No. 336-1. Plaintiffs need the SFT data to confront that defense; the very need for these late-stage tools in the first place bears on lack of “transformation.” RFP 118 squarely covers SFT data, and it is also grounded in the Llama Papers, as referenced above.

Databases. Meta never contends the EAG-WSF hard drive clusters are irrelevant, nor does Meta adequately explain the burden of producing these datasets. Meta raises numerous procedural arguments, but the EAG-WSF clusters clearly must contain the data that Plaintiffs seek; otherwise, Meta would have merely stated it had no such responsive data. Similarly, Meta does not refute the fact that there is a not-yet-produced “CRS 17 – Intellectual Property” database referenced within recently produced source code. At bottom, Plaintiffs seek narrow relief to obtain essential post-training data that Meta doesn’t dispute exists and relies on to support its fair use defense.

By: /s/ Bobby Ghajar

Bobby A. Ghajar
Colette A. Ghazarian
COOLEY LLP
1333 2nd Street, Suite 400
Santa Monica, CA 90401
Telephone: (310) 883-6400
Facsimile: (310) 883-6500
Email: bghajar@cooley.com
cghazarian@cooley.com

Mark R. Weinstein
Elizabeth L. Stameshkin
COOLEY LLP
3175 Hanover Street
Palo Alto, CA 94304
Telephone: (650) 843-5000
Facsimile: (650) 849-7400
Email: mweinstein@cooley.com
lstameshkin@cooley.com

Kathleen R. Hartnett
Judd D. Lauter
COOLEY LLP
3 Embarcadero Center, 20th Floor
San Francisco, CA 94111
Telephone: (415) 693-2071
Facsimile: (415) 693-2222
Email: khartnett@cooley.com
jlauter@cooley.com

Phillip Morton
COOLEY LLP
1299 Pennsylvania Avenue, NW, Suite 700
Washington, DC 20004
Telephone: (202) 842-7800
Facsimile: (202) 842-7899
Email: pmorton@cooley.com

Angela L. Dunning
**CLEARY GOTTlieb STEEN &
HAMILTON LLP**
1841 Page Mill Road, Suite 250
Palo Alto, CA 94304
Telephone: (650) 815-4121

By: /s/ Maxwell V. Pritt

BOIES SCHILLER FLEXNER LLP
David Boies (*pro hac vice*)
333 Main Street
Armonk, NY 10504
(914) 749-8200
dboies@bsfllp.com

Maxwell V. Pritt (SBN 253155)
Joshua M. Stein (SBN 298856)
44 Montgomery Street, 41st Floor
San Francisco, CA 94104
(415) 293-6800
mpritt@bsfllp.com
jstein@bsfllp.com

Jesse Panuccio (*pro hac vice*)
Jay Schuffenhauer (*pro hac vice*)
1401 New York Ave, NW
Washington, DC 20005
(202) 237-2727
jpanuccio@bsfllp.com
jschuffenhauer@bsfllp.com

Joshua I. Schiller (SBN 330653)
David L. Simons (*pro hac vice*)
55 Hudson Yards, 20th Floor
New York, NY 10001
(914) 749-8200
dsimons@bsfllp.com
jischiller@bsfllp.com

Interim Lead Counsel for Plaintiffs

Facsimile: (650) 849-7400
Email: adunning@cgsh.com

Attorneys for Defendant Meta Platforms, Inc.

ATTESTATION PURSUANT TO CIVIL LOCAL RULE 5-1(h)

I hereby attest that I obtained concurrence in the filing of this document from each of the other signatories. I declare under penalty of perjury that the foregoing is true and correct.

Dated: December 23, 2024

BOIES SCHILLER FLEXNER LLP

/s/ Maxwell V. Pritt

Maxwell V. Pritt

Reed Forbush

Jay Schuffenhauer

Attorneys for Plaintiffs